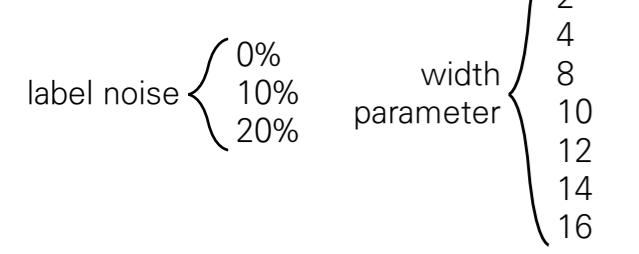


1. ResNet18

1) CIFAR-100

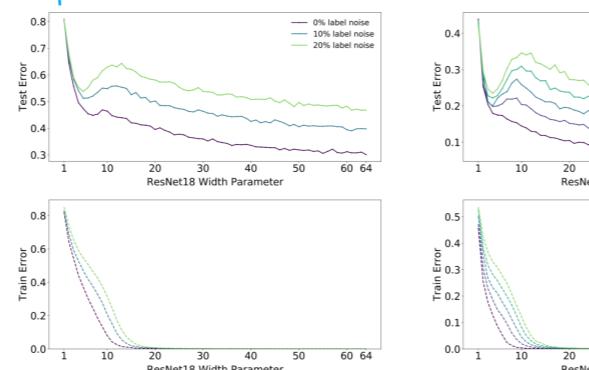


Adam,
lr = 0.0001
1000 epochs
(논문은 4000 epochs)

성대규

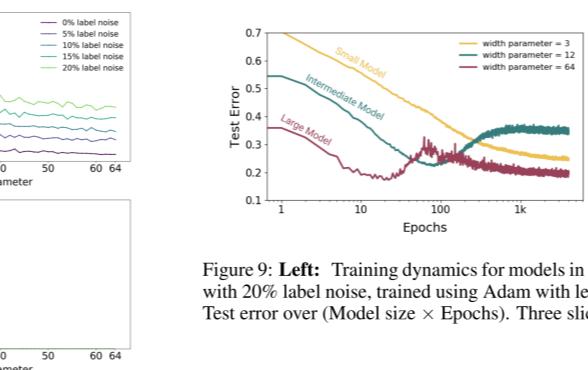
$3 \times 8 = 24$
완료됨

결과



(a) CIFAR-100. There is a peak in test error even with no label noise.

결과

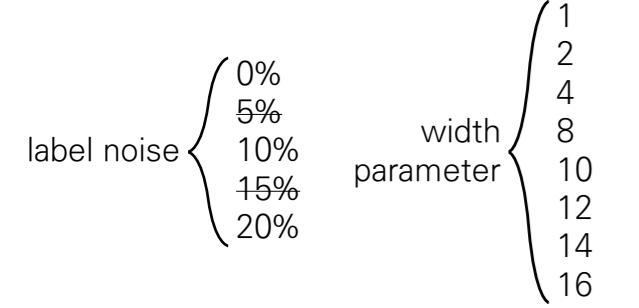


(b) CIFAR-10. There is a "plateau" in test error around the interpolation point with no label noise, which develops into a peak for added label noise.

Figure 4: Model-wise double descent for ResNet18s. Trained on CIFAR-100 and CIFAR-10, with varying label noise. Optimized using Adam with LR 0.0001 for 4K epochs, and data-augmentation.

Fig 4 (5페이지): (a)는 세연님, (b)는 지윤님

2) CIFAR-10

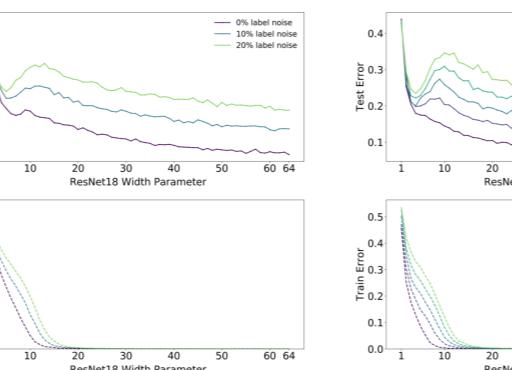


Adam,
lr = 0.0001
1000 epochs
(논문은 4000 epochs)

홍세연 님,
전지윤 님

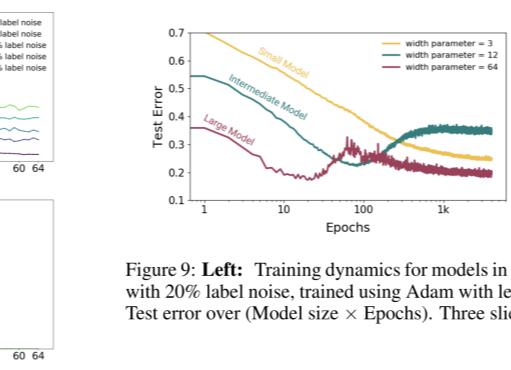
$3 \times 8 = 24$
(논문은 $5 \times 7 = 35$)

결과



(a) CIFAR-100. There is a peak in test error even with no label noise.

결과



(b) CIFAR-10. There is a "plateau" in test error around the interpolation point with no label noise, which develops into a peak for added label noise.

Figure 4: Model-wise double descent for ResNet18s. Trained on CIFAR-100 and CIFAR-10, with varying label noise. Optimized using Adam with LR 0.0001 for 4K epochs, and data-augmentation.

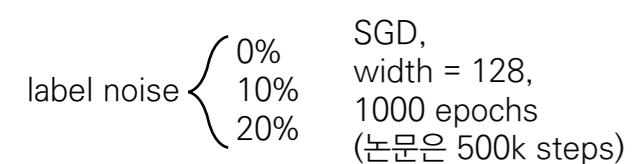
Fig 4 (5페이지): (a)는 세연님, (b)는 지윤님

2. 5-layer CNN

1) CIFAR-10

(Adam의 경우 필요한 width parameter: 1, 2, 4, 8, 16, 32, 64 (Fig 6)) 완료됨

optimizer { Adam: 1000 epochs (논문은 4000 epochs)
SGD: 1000 epochs (논문은 500k steps)

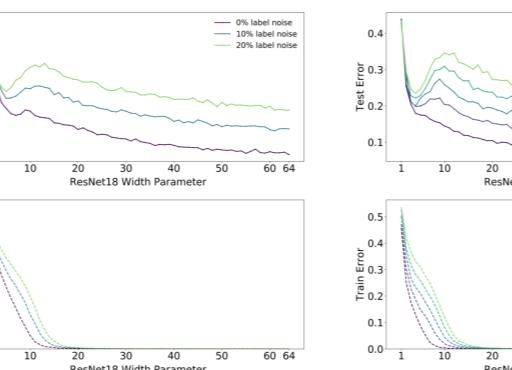


SGD,
width = 128,
1000 epochs
(논문은 500k steps)

성대규

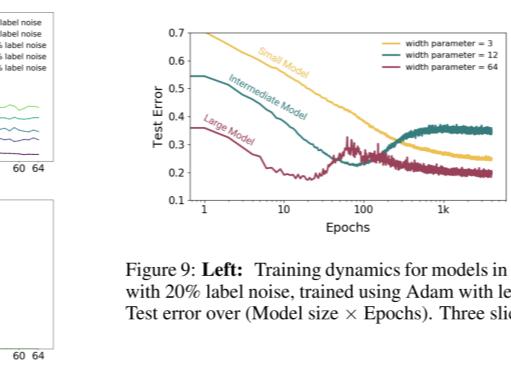
$3 \times 1 = 3$ 완료됨

결과



(a) CIFAR-100. There is a peak in test error even with no label noise.

결과



(b) CIFAR-10. There is a "plateau" in test error around the interpolation point with no label noise, which develops into a peak for added label noise.

Figure 4: Model-wise double descent for ResNet18s. Trained on CIFAR-100 and CIFAR-10, with varying label noise. Optimized using Adam with LR 0.0001 for 4K epochs, and data-augmentation.

Fig 4 (5페이지): (a)는 세연님, (b)는 지윤님

3. Transformer

dataset { IWSLT
WMT

시간이...?

최우선) 간단한 거로 D.D 현상이 보이는지 확인
그 후 가능한 만큼 실험 돌리기
train, test epoch별 모두 기록

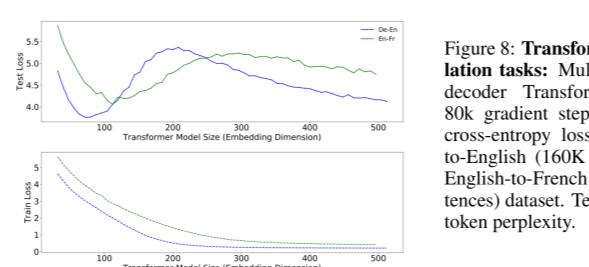
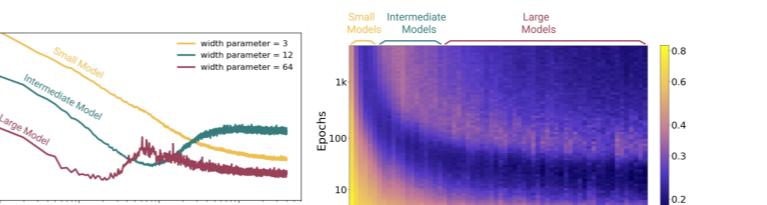


Fig 8 (7페이지): 기윤님

Figure 8: Transformers on language translation tasks: Multi-head-attention encoder-decoder Transformer model trained for 80k gradient steps with labeled smoothed cross-entropy loss on IWSLT'14 German-to-English (160K sentences) and WMT'14 English-to-French (subsampled to 200K sentences) dataset. Test loss is measured as per-token perplexity.



Small Models
Intermediate Models
Large Models

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8

1k

Epochs

ResNet18 Width Parameter

Test Error

Colorbar

0.2
0.4
0.6
0.8